EXPLORING HIGH SCHOOL STUDENTS' CARELESS OR INSUFFICIENT EFFORT SURVEY RESPONSES WHEN INVESTIGATING ATTITUDES TOWARD MATHEMATICS

Xiaohui Wang
The University of Texas Rio Grande Valley
xiaohui.wang@utrgv.edu

Mayra Ortiz Galarza
The University of Texas Rio Grande Valley
mayra.ortizgalarza@utrgv.edu

Sergey Grigorian
The University of Texas Rio Grande Valley
sergey.grigorian@utrgv.edu

Aaron Wilson
The University of Texas Rio Grande Valley
aaron.wilson@utrgv.edu

John Knight
The University of Texas Rio Grande Valley
john.knight01@utrgv.edu

A modified, bilingual Attitudes Toward Mathematics Inventory (ATMI) instrument was administered to 1,258 high school students in South Texas in an NSF-funded project on informal learning of mathematics and near peer mentoring. We explore students' survey response behaviors and examine the existence of careless and insufficient effort (CIE) responses. This is empirical research for handling the challenge of CIE responses that leads to improved survey data quality, thus eventually validating the intervention effect of the mathematical informal learning project.

Keywords: high school education; informal education; affect, emotion, beliefs and attitudes; measurement

It is well-known that in educational and psychological research self-report measures (i.e., surveys, questionnaires, inventories, etc.) are commonly used by researchers and practitioners to measure non-cognitive constructs (Ulitzsch, Yildirim-Erbasli, Gorgun, & Bulut, 2022). However, in these measures, Careless Response (CR) and Insufficient Effort Response (IER) have been prevalently identified as common sources for biased estimates, and invalid inferences and response data (see for example, Zhong, Li, & Li, 2021; Ward & Meade, 2023).

As part of an NSF-funded project on informal learning of mathematics, the authors implemented an informal educational intervention in which high school students interacted with their near-peer college students in the context of novel mathematical explorations (Wilson, et al., 2023). In particular, high school students were exposed to *MathShows,* which were interactive mathematical presentations, as well as math social media, and a math summer internship. A smaller scale pilot study (Wilson & Grigorian 2019) has shown that such near peer interventions have the potential to positively affect attitudes to mathematics. To understand the impact of such interactions on high school students' attitudes to mathematics, the project team administered surveys to the high school students involved in the project before the start of the project, at specific times during the project, and at the end of the project. The experimental design involved some schools receiving the intervention, while others serving as control groups with no intervention (i.e., intervention offered after data collection). Over the course of 3 years, this project involved 1,258 students from 4 high schools in two majority-Hispanic school districts in South Texas. One of the survey instruments was the 19-item shortened Attitudes Toward Mathematics Inventory (ATMI) survey (Lim & Chapman, 2013; Majeed et al., 2013, Wilson & Grigorian 2019). This is a 5-point

Likert scale survey that measures four subscales: mathematics enjoyment, mathematics motivation, self-confidence in mathematics, and perceived value of mathematics.

Measuring beliefs and attitudes toward mathematics is important to mathematics educational research. In this empirical study, we address a necessary concern about validity of responses in any survey-based educational research project.

Research Context and Methods of the Study

A qualitative component of the study involved focus group interviews with high school students. Reported data derived from some focus groups show that some participants answered the surveys carelessly. Some of the careless survey completion behavior includes answering questions without reading them, choosing answers at random, and copying answers from their classmates. This anecdotal evidence of CIE (careless or insufficient effort) responses prompted concern on how widespread this practice may be, especially since some surveys were administered remotely during the COVID-19 pandemic without project team supervision. Therefore, to accurately measure the impact of the intervention activities on student attitudes toward mathematics, it has become necessary to implement robust techniques to detect and filter CIE responses. C/IE responses can evidently invalidate research results since they represent data that is derived from a scenario where the survey respondent failed to provide genuine responses (Curran, 2015). Throughout the paper we use the acronym CIE for *careless and/or insufficient effort survey responses*: "in which a person responds to items without sufficient regard to the content of the items and/or survey instructions" (Huang, Liu, & Bowling, 2015, p. 828).

In recent years, multiple methods for detecting both CR and IER have been developed and amended to better detect a wide array of CR and IER patterns indicators (i.e., response-pattern or response-time-based, model-based approaches, etc.) (Ulitzsch, et al., 2022). For instance, Curran (2015) reviews a variety of methods that have been used to detect survey responses that were evidently provided by participants who were careless or else gave insufficient effort C/IE. For detection of such responses, Curran suggests applying certain methods in sequence, as deemed appropriate to the circumstances of the survey administration and data received (Curran, 2015). Multiple post-hoc metrics or methods for the detection of C/IE responses are highlighted by Curran, including: 1) response time, 2) long-string analysis, 3) Mahalanobis distance, 4) odd-even consistency, 5) resampled individual reliability, 6) semantic antonyms/synonyms, 7) inter-item standard deviation, 8) polytomous Guttman errors, and 9) person total correlation. Beyond these, Curran (2015) also suggests other checks that can be included when developing the survey, before data collection. For instance, of particular interest to our study is Curran's observation that surveys involving reverse-worded items introduce additional complexity into the detection of C/IE responses. In this paper we address data from a survey that included some reverse-worded items, and thus our findings elaborate the detection of C/IE responses for this more complex case.

Research Objectives and Data

Our research goal is to investigate to what extent high school students make CIE responses when taking the ATMI survey. Table 1 lists the 19 ATMI items used in our study, with the tone of negative or positive marked on the side. The two research questions we will address are:

- What are the CIE indicators for this study?
- To what extent did high school students make CIE responses?

We use the baseline data collected digitally via Qualtrics surveys from 2021 through 2022. Figure 1 gives a sample view of how the survey appeared to students. Students' participation was voluntary. The data consists of survey responses and time stamp information from 1,258 students from four different schools (A, B, C and D). The first two schools (A intervention, and B control) were in the remote schooling format (due to COVID pandemic) during the data collection stage, thus surveys had to be supervised by teachers. The other two schools (C and D) were in the face-to-face schooling format when data was collected. The survey was administered by the research team in the school classrooms. It is worth noting that two different administration methods were used among the latter two schools. In School C, an intervention school, students were given the survey including consent and assent forms right after the introduction of the project and research team personnel. In School D, a control school, a short motivating *mini-MathShow* (a "math-magician" show involving a number trick) was performed after those introductions and before the survey administration, in an effort to build up an initial rapport between the project team and the students. During the surveys administration in Schools C and D, the researchers persistently reminded students about the requirement and importance of reading the survey questions carefully and responding accurately and honestly.

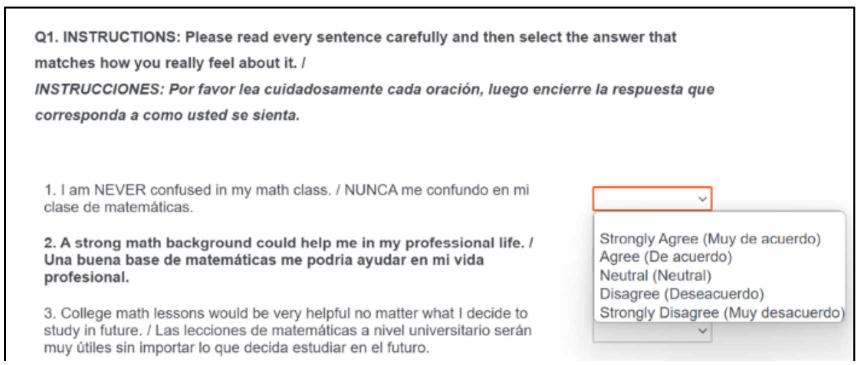

Figure 1: Sample View of The Online ATMI Survey

The table below gives all 19 items of ATMI as implemented in this study, noting the working orientation (positive or negative), as well as providing the Spanish translation as used in our administration.

Table 1: The 19 ATMI Items Used (Four constructs: SC – Self-Confidence; VAL – Sense of Value Toward Math, ENJ – Enjoyment of Math, MOV – Motivation to Do Math)

Tone and Construct	ATMI Items
+ SC	1. I am never confused in my math class. (Nunca me confundo en mi clase de matemáticas)
+ VAL	2. A strong math background could help me in my professional life. (Una buena base de matemáticas me podria ayudar en mi vida profesional)
+ VAL	3. College math lessons would be very helpful no matter what I decide to study in future. (Las lecciones de matemáticas a nivel universitario serán muy útiles sin importar lo que decida estudiar en el futuro)
- ENJ	4. Math is NOT a very interesting subject. (Las matemáticas NO son una materia muy interesante.)
+ ENJ	5. I really like math. (Me gustan mucho las matemáticas)
- SC	6. It makes me nervous to even think about having to do a math problem. (Me pone nervioso tan solo pensar en hacer un problema de matemáticas)
- ENJ	7. I don't like to solve new problems in math. (No me gusta resolver problemas nuevos de matemáticas.
- VAL	8. Math is one of the LEAST important subjects for people to study. (Las matemáticas son una de las materias MENOS importantes que la gente debe estudiar.)
- SC	9. I feel a sense of insecurity when attempting math. (Siento una sensación de inseguridad cuando intento hacer matemáticas.)
- VAL	10. Math is a worthless and unnecessary subject. (Matemáticas es una materia sin valor e innecesaria.)
- MOV	11. The challenge of math does not appeal to me. (El reto de hacer matemáticas no me llama la atención.)
+ ENJ	12. I am happier in a math class than in any other class. (Soy más feliz en una clase de matemáticas que en cualquier otra clase.)
- SC	13. Studying math makes me feel nervous. (Estudiar matemáticas me hace sentir nervioso.)
- MOV	14. I plan to take as little math as I can during my education. (Yo planeo tomar las menos matemáticas possibles durante mi educación.)
+ ENJ	15. I have usually enjoyed studying math in school. (Usualmente he disfrutado estudiar matemáticas en la escuela.)
- VAL	16. Math is NOT important in everyday life. (Las matemáticas NO son importantes en la vida diaria.)
+ SC	17. I am always calm and relaxed in a math class. (Siempre estoy calmado y relajado en una clase de matemáticas.)
+ MOV	18. I am willing to take more than the required amount of math. (Estoy dispuesto a tomar más matemáticas de lo requerido.)
+ MOV	19. I am confident that I could learn advanced math. (Me siento seguro de que podría aprender matemáticas avanzadas.)

CIE Identifying Methods

We coded the 5-point Likert scales for the ATMI survey items as shown in Figure 1 and Table 1 using values of (-2, -1, 0, 1, 2). To distinguish the severeness of the CIE responses, we labeled them in four categories: non-CIE, slight CIE, moderate CIE, and severe CIE. We assigned flag value of 0 to both non-CIE and slight CIE cases, value 1 to moderate CIE, and value 2 to severe CIE. Combining the methods existing in the literature and our survey, we propose the following criteria to identify potential CIE responses:

1. *All-item-same criterion*: if a student's responses for all 19 items are the same, the responses will be classified as severe CIE. Flag values for this criterion are 0 or 1, with 1 representing CIE flag.

2. *Response time criterion*: Abandoned or non-submitted surveys are automatically recorded by Qualtrics after a period of 7 days, so existence of such responses leads to extreme outliers in survey duration. On the other hand, testing of the online surveys by the research team has shown that it is not feasible to complete the entire survey too quickly, while reading the questions and answering thoughtfully. If a student's survey duration (recorded by Qualtrics) is shorter than the threshold, defined by "median – median absolute deviation" that is calculated based on all survey durations in his/her school. Median and median absolute deviation are used for threshold computing because the response time data has many far outliers. This is the total response time, not just the ATMI, which was only one component of a longer survey. Flag values for this criterion is 0 or 1, with 1 representing CIE flag.
3. *ATMI score outliers:* if a student's ATMI score appeared to be an outlier compared to the ATMI scores in the corresponding school, we classify it as a potential CIE.
4. *Neighbor Opposite-Item-Pair consistency criterion*: For a pair of two neighbor items, if one is positively toned and the other negatively toned, it is a Neighbor Opposite-Item-Pair. There are three such pairs as shown in Table 1, and they are items 4&5, 11&12, and 14&15. Out of the 25 possible responses combinations of the choices for an opposite pair, if the two codes had different signs or being zero in one of the items, we consider that the student had polarized answers, which indicates that she/he reacted to the tone switch correctly. These types of combinations are classified as non-CIE. The combinations (-1)to(-1) or (1)to(1) fall outside of the class of non-CIE but since the inconsistency is minor, we classify it as slight CIE. We classify (-2)to(-2) or (2)to(2) as severe CIE for these choices pair indicating maximum possible inconsistency. Four other combinations are classified as moderate CIE.
5. *Same Construct Opposite-Item-Pair consistency criterion*: For a pair of two items in a construct, if one is positively toned and the other negatively toned and that expert opinion confirms a high level of similarity among them, it is a Same Construct Opposite-Item-Pair. There are four such pairs as shown in Table 1: three pairs of SC construct items, 6&17, 9&17, and 13&17; and one pair of ENJ construct items, 4&5. This last pair is already included in the Neighbor Opposite-Item-Pair consistency criterion. The same classification methods as in Neighbor Opposite-Item-Pair consistency criterion are used for pairs in this criterion.
6. *Same Construct Similar-Item-Pair consistency criterion*: For a pair of two items in a construct, if they are toned in the same direction and that an expert opinion confirms a high level of similarity among them, it is a Same Construct Similar-Item-Pair. There are eight such pairs as shown in Table 1: three pairs of SC construct items,6&9, 6&13, and 9&13; one pair of ENJ construct items: 12&15; and four pairs of VAL construct items: 2&3, 8&10, 8&16, and 10&16.There are maximum of four units steps between the choices, from (–2)to(2) or (2)to(–2). We classify any response combinations having one step apart or completely agree is non-CIE. If the two choices are two steps apart, it is a slight CIE; three steps apart, moderate CIE, and then four steps apart, severe CIE.

All the aforementioned criteria are constructed in a conservative way to allow students to express their opinions and to avoid disqualifying responses that have limited accidental errors.

Results

Of the 1,258 student participants, we found that there were 19 students (1.5%) giving the same answers to all 19 ATMI items. According to the response time criterion, there were 90 students (7.8%) flagged for responding too fast. Figure 2 (Panel a) displays the distribution of survey durations for four schools. The distributions of survey durations varied. One reason is that the total numbers of survey items varied (51, 51, 36 and 47 items for Schools A, B, C and D). School D seemed to have relatively longer survey durations than other schools. According to the ATMI score outliers criterion, there were 20 students (1.6%) flagged for their extreme ATMI scores. Of them, nine students had low ATMI scores (≤34) and eleven students reported high ATMI scores (≥87). Figure 2 (Panel b) displays the distribution of ATMI scores for four schools. We see that the four schools have almost identical distributions of ATMI scores at baseline. Note in Figure 2 that the y-axis for survey duration was cut-off at 40 mins, as there were a number of extremely far outliers, indicative of students that forgot to submit their form until hours and days later.

Table 2 lists the students' response pairs and corresponding counts for the Opposite-Item-Pairs. Those pairs are grouped into four levels of CIE. We found that more than 84% of students gave consistent responses to those six pairs, respectively. Students' CIE rates for the Neighbor Opposite-Item-Pairs increased for questions later in the survey, from 5.3% on the first pair of 4&5, to 12.3% on the second pair of 11&12, then to 16% on the last pair 14&15. The rates of the three CIE sub-categories also increased: slight CIE increased from 3.4% to 8.4% and then to 10.1%; moderate CIE increased from 1.4% to 3.6% and then to 4.5%; severe CIE increased from 0.6% to 1.2% then to 1.4%.

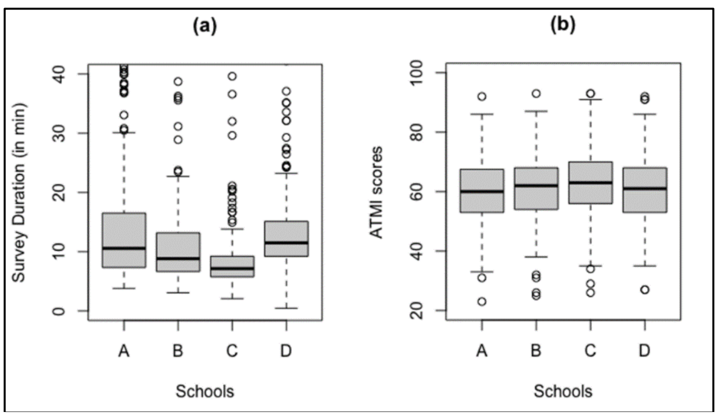

Figure 2: The Distributions of Survey Durations (Panel a) and ATMI Scores (Panel b) for Four Schools.

Table 2. Students' Opposite-Item-Pair Responses in Different CIE Levels

Pair-wise Responses	Neighbor Opposite-Item-Pair			Same Construct Opposite-Item-Pair (SC)		
	4&5	11&12	14&15	6&17	9&17	13&17
9 pairs with one or two 0s	648	825	678	676	652	629
(-2)to(1)	30	21	30	41	43	37
(-2)to(2)	55	13	20	29	32	32
(-1)to(1)	203	68	166	186	175	203
(-1)to(2)	36	7	39	34	24	30
(1)to(-2)	40	43	8	18	22	20
(1)to(-1)	83	72	70	81	101	107
(2)to(-2)	61	28	21	32	28	30
(2)to(-1)	30	13	24	34	31	19
Non-CIE	1186 (94.7%)	1090 (86.7%)	1056 (84.0%)	1131 (90.0%)	1108 (88.1%)	1107 (88.1%)
(-1)to(-1)	20	94	34	38	45	44
(1)to(1)	22	11	93	44	63	61
Slight CIE	42 (3.4%)	105 (8.4%)	127 (10.1%)	82 (6.5%)	108 (8.6%)	105 (8.4%)
(-2)to(-1)	2	8	3	3	1	4
(-1)to(-2)	10	31	9	8	11	11
(1)to(2)	3	6	30	14	9	15
(2)to(1)	3	2	15	11	7	6
Moderate CIE	18 (1.4%)	47 (3.7%)	57 (4.5%)	36 (2.9%)	28 (2.2%)	36 (2.9%)
(-2)to(-2)	4	9	5	4	5	4
(2)to(2)	3	6	12	4	8	5
Severe CIE	7 (0.6%)	15 (1.2%)	17 (1.4%)	8 (0.6%)	13 (1.0%)	9 (0.7%)
Total	1253	1257	1257	1257	1257	1257

The Similar-Item-Pairs counts for all possible response pairs and three levels of CIE are reported in Table 3. We found that almost more than 90% of students gave consistent responses on those eight pairs, respectively. The pairs involving later items in the survey had higher CIE rates than pairs involving earlier items.

Table 3. Students' Similar-Item-Pair Responses in Different CIE Levels

Constructs	SC (3 pairs)			ENJ (1 pair)	VAL (4 pairs)			
Pair-wise Responses	6&9	6&13	9&13	12&15	2&3	8&10	8&16	10&16
Non-CIE (15 pairs with values less than 2 units apart)	1196 (95.2%)	1209 (96.3%)	1197 (95.3%)	1136 (90.3%)	1178 (93.9%)	1193 (95.0%)	1148 (91.3%)	1150 (91.5%)
(-2)to(0)	9	3	4	35	4	4	38	57
(0)to(-2)	9	8	13	2	5	25	19	4
(0)to(2)	12	7	8	47	18	5	12	7
(2)to(0)	12	11	15	3	25	9	8	3
Slight CIE	42 (3.3%)	29 (2.3%)	40 (3.2%)	87 (6.9%)	52 (4.1%)	43 (3.4%)	77 (6.1%)	71 (5.6%)
(-2)to(1)	1	2	4	15	3	1	6	16
(-1)to(2)	6	2	5	8	7	2	7	10
(1)to(-2)	1	5	2	4	6	8	5	1
(2)to(-1)	5	6	5	2	3	4	4	3
Moderate CIE	13 (1.0%)	15 (1.2%)	16 (1.3%)	29 (2.3%)	19 (1.5%)	15 (1.2%)	22 (1.8%)	30 (2.4%)
(-2)to(2)	2	2	2	5	3	0	4	4
(2)to(-2)	3	1	1	1	2	5	6	2
Severe CIE	5 (0.4%)	3 (0.2%)	3 (0.2%)	6 (0.5%)	5 (0.4%)	5 (0.4%)	10 (0.8%)	6 (0.5%)
Total	1256	1256	1256	1258	1254	1256	1257	1257

Based on the six criteria, the maximum possible total count of flags is 17. The more flags a response has, the more likely it is a CIE response. Table 4 gives the distribution of flag counts for the 1258 student participants. There were 390 (31%) students having at least one CIE flag, 135 (10.7%) students having at least two CIE flags, and 57 (4.5%) students having at least three CIE flags.

Table 4. Summary of Students with Various Number of Flags

No. of Flags	7	6	5	4	3	2	1	0
Counts (%)	2(0.2%)	3 (0.2%)	5 (0.4%)	13 (1%)	34 (2.7%)	78 (6.2%)	255 (20.3%)	868 (70%)
Cumulative Counts (%)	2 (0.2%)	5 (0.4%)	10 (0.8%)	23 (1.8%)	57 (4.5%)	135 (10.7%)	390 (31%)	1258 (100%)

Discussion

Some methods for identifying CIE responses as found in the literature (e.g., Curran, 2015). are not suitable for our digital survey situation. For example, the long-string or straight-lining criterion does not seem to be relevant for the survey administered in this study, because participants needed

to select responses from drop-down boxes, as shown in Figure 1. Moreover, long strings may be genuine responses for some items in our survey (such as items 6 through 11 could give long-string = 6). The digital survey platform, Qualtrics, allowed us to record the survey duration. We used survey questions with mixed tones consisting of ten negative tone items and nine positive tone items. This mixture gives us a total of 14 pairs to examine consistency of participants' responses. With such survey questionnaire design, we were able to propose the six criteria to identify CIE responses for the study population, leading to our answer for the first research question.

For the second research question, there were significant number of students who gave CIE responses as close to a third of participating students had at least one CIE flag. As noted in literature (Curran, 2015; Zhong et al., 2021) the rate of CIE responses increases as the survey progresses, and we have observed the same pattern with CIE rates for the Neighbor Opposite-Item-Pairs increasing for later questions in the survey.

Conclusion

The findings in this study confirm the existence of CIE responses. In the next phases of this study, it will be essential to examine how CIE responses impact the ATMI reliability measures and the corresponding confirmatory factor analysis. Further, we will examine whether different survey administration methods (remote vs face-to-face schooling formats) affect the occurrence of CIE responses. The surveys analyzed for this paper are the baseline surveys at the beginning of the project for each school. The study reported here is ongoing and we will soon collect end-of-project (EOP) surveys. The CIE criteria calibrated from the baseline surveys will be applied to the EOP surveys. It is worth noting that calibration of CIE criteria should not be based on EOP surveys to avoid bias.

Using the methods detailed in this paper allows filtering CIE responses to measure the differences more accurately in math attitudes between control and intervention schools. The survey design strategy and CIE detection methods are generalizable to research aiming at learner's belief and attitude evaluations, especially study involving high school students.

References

Curran, P. G. (2015). Methods for the detection of carelessly invalid responses in survey data. Journal of Experimental Social Psychology, 66 (2016), 4–19. https://doi.org/10.1016/j.jesp.2015.07.006

Huang, J. L., Bowling, N. A., Liu, M., & Li, Y. (2015). Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions. Journal of Business and Psychology, 30(2), 299–311. https://doi.org/10.1007/s10869-014-9357-6

Lim, S. Y., Chapman, E. (2013). Development of a short form of the attitudes toward mathematics inventory. Educational Studies in Mathematics, 82(1), 145–164.

Majeed, A. A., Darmawan, I. G. N., Lynch, P. (2013). A confirmatory factor analysis of attitudes toward mathematics inventory. The Mathematics Educator, 15(1), 121–135.

Ulitzsch, E., Yildirim-Erbasli, S. N., Gorgun, G., & Bulut, O. (2022). An explanatory mixture IRT model for careless and insufficient effort responding in self-report measures. British Journal of Mathematical and Statistical Psychology, 75(3), 668–698. https://doi.org/10.1111/bmsp.12272

Ward, M. K., & Meade, A. W. (2023). Dealing with careless responding in survey data: Prevention, identification, and recommended best practices. Annual Review of Psychology, 74, 577–596. https://doi.org/10.1146/annurev-psych-040422-045007

Wilson, A. T., & Grigorian, S. (2018). The near-peer mathematical mentoring cycle: Studying the impact of outreach on high school students' attitudes toward mathematics. International Journal of Mathematical Education in Science and Technology, 50(1), 46–64. https://doi.org/10.1080/0020739X.2018.1467508

Wilson, A. T., Wang, X., Galarza, M. O., Knight, J., & Patino, E. (2023). Math attitudes and identity of high schoolers impacted through participating in informal, near-peer mentoring. International Journal of Research in Education and Science (IJRES), 9(2), 535-545. https://doi.org/10.46328/ijres.3093

Zhong, X., Li, M., & Li, L. (2021). Preventing and detecting insufficient effort survey responding. Advances in Psychological Science, 29(2), 225–237. https://doi.org/10.3724/SP.J.1042.2021.00225